This full-text paper was peer-reviewed and accepted to be presented at the IEEE WiSPNET 2016 conference.

Review of Techniques used for Sensory Data Analytics Over Cloud

Manujakshi B C
Dept of Computer Science & Engg
Acharya Institute of Technology
Bangalore, India
E-Mail: manujakshibc2014@gmail.com

K B Ramesh
Dept. of Electronics & Instrumentation Engg.
RVCE, Bangalore, India

***Abstract*: - In recent years, there is a high growth in wireless sensor network (WSNs) for secure communications or sensor network applications. It has been seen that sensor applications are revised with upcoming technologies in order to meet the demands. There is a requirement of efficient design and implementation of WSNs, due to the huge SNs to allow applications which connect the physical world to the virtual world. With a wide range of applications for SNs, some of the application areas are health, military, and security will create issues in data transmissions from one sensor network to another sensor network. As sensors gathers complex data, it is quite a difficult job to understand how far it can be valuable with respect to analysis. Owing to inherent complexities e.g. size, heterogeneity, un-structuredness, etc. it may pose serious problems in futuristic sensor data analytics over cloud. This paper discusses about the techniques used in managing sensor data over cloud to understand the existing system and its effectiveness.**

Keywords- Data Analytics on the Cloud, Sensor Network, Secure Communications, Wireless Sensor Network.

I. INTRODUCTION

With the modernization as well as ubiquitousness of the commercial application, there is a huge generation of data. The modern sensor-based applications e.g. IoT are tremendously increasing that collectively generates an exponential data [1]. Such data is also subjected to the storage systems of data centers across a broad range of industries. Sensor are being used in various application e.g. the military application, battlefield surveillance, area monitoring, commercial application at home and industries, event detection (intrusion/flood emergencies) etc. The price of sensor nodes is quite variable depending on the complication to be handled by the individual sensor nodes. Dimension and price constraints on sensor nodes result in equivalent constraints on resources such as energy, memory, computational speed, and communications bandwidth [2]. The topology of the WSNs is able to differentiate from a straightforward star network to a superior multi-hop wireless mesh network. The propagation method between the hops of the network can be either routing based or peer-to-peer based. The unique characteristic of the sensor network; however, also bring various problem and challenges, the specified by sensor data collection applications also increase issues that require to be considered in the network structure [3]. First of all, the installed sensors might need to cover up the full region that the sensor data collecting application is interested. Secondly, when two or newly added sensors send data to a common neighbor at the similar time, the data smash together at the common neighbor that will not obtain any of these data. In order to gain data accurately, sensors may have needed to be set at specific locations. Various types of data can be acquired by different sensors (motions, vibration, and temperature, light) with different sampling rates [4]. This problem may cause disturbed power consumption over a wireless sensor network. As well as reduce the network lifetime if not managing useful way and since data are to be sent. To the main base network without corrupt and lacking any data the data aggregation/fusion procedure are tough to be applied [5]. The enormous quantity of data send by the sensor network requires a massive storage and computing infrastructure to process and analyze the sensor data. Hence, the "Service Oriented Sensor Network Architecture" (SOSA) is extended to cloud architecture through Integration controller, in which this sensor services is deployed into a public cloud. Big data analytics is a set of advanced technologies to work with large volumes (Intel, 2014) of heterogeneous data [6]. Hence, it is chosen as the technology to analyze the sensor data, which are coming from heterogeneous SN, such as temperature SNs, pressure SNs, humidity SNs, vision SNs, etc. The data formats from different SNs are physical quantities such as degree centigrade, degree Kelvin, psi, kg/m^2, etc., the data are classified as unstructured data. Unstructured data will be converted into a unified data format, i.e., XML, which is categorized as semi-structured data. This XML conversion is necessary to change the sensor data into the web service message. It is essential to integrate WSN with SOA and then extend to the cloud. XML is used to carry the sensor data but at the same time, it is the protocol language of the Internet which can communicate with systems on any platform, and any architecture. Section II has presents Techniques of sensor data analysis on cloud and research work followed in Section III. The Application of sensors and system illustrates in Section IV followed by Result gap has followed in V. Finally, summary of the paper is carried out in Section VI.

II. TECHNIQUES OF SENOR DATA ANALYSIS ON CLOUD

The standard techniques of sensor data analysis in cloud are as follows:

978-1-4673-9338-6/16/$31.00 ©2016 IEEE

A. Classification:-

Characterization Classification is the most normally connected data mining procedure, which utilizes an arrangement of pre-ordered illustrations to add to a model that can group the number of inhabitants in records on the loose. Misrepresentation recognition and credit hazard applications are especially appropriate to this sort of investigation. This methodology as often as possible utilizes choice tree or neural system based characterization calculations. The data arrangement procedure includes learning and order. In learning characterization calculation investigates the preparation data [7]. In arrangement test data are utilized to assess the exactness of the order rules. On the off chance that the precision is adequate, the guidelines can be connected to the new data tuples. For an extortion identification application, this would incorporate complete records of both deceitful, and legitimate exercises decided on a record-by-record premise. The classifier-preparing calculation utilizes these pre-grouped cases to focus the arrangement of parameters needed for fitting separation. The calculation then encodes these parameters into a model called a classifier.

B. Data Senor Clustering:-

By utilizing clustering systems, we can further distinguish thick and inadequate locales in objects space and can find general conveyance example and connections among information traits. Order methodology can likewise be utilized for a successful method for recognizing gatherings or classes of objects yet it turns out to be expensive so clustering can be utilized as a preprocessing methodology for quality subset choice and arrangement. Case in point, to frame gathering of clients based on purchasing patterns obtaining examples, to classifications qualities with comparative functionality.

C. Regression:-

Regression procedure can be adjusted for prediction. Relapse investigation can be utilized to show the relationship between one or more free variables and ward variables. In data mining, free variables are characteristics known, and reaction variables are what we need to anticipate. Shockingly, some true issues are not just forecast. For example, deals volumes, stock costs, and item disappointment rates are all extremely hard to anticipate because they may rely on upon complex communications of different indicator variables. In this way, more mind boggling systems (e.g., logistic relapse, choice trees, or neural nets) may be important to figure future qualities. The same model sorts can regularly be utilized for both relapse and characterization.Case in point, the CART (Classification and Regression Trees) choice tree calculation can be utilized to assemble both grouping trees (to arrange absolute reaction variables) and relapse trees (to conjecture consistent reaction variables). Neural systems also can make both characterization and regression models.

D. Association Rule:-

Association and correlation are normal to discover incessant thing set discoveries among expansive data sets. These kind of discovering helps organizations to settle on specific choices, for example, list outline, cross advertising and client shopping conduct examination. Affiliation Rule calculations should have the capacity to create rules with certainty values under one. However the quantity of conceivable Association Rules for a given dataset is by, and large huge and a high extent of the standards are for the most of little (if any) quality.

E. Neural network:-

The neural network is a situated of joined info/yield units, and every association has a weight present with it. Amid the learning stage, the system learns by changing weights to have the capacity to foresee the right class names of the info tulles [7]. Neural systems have the astounding capacity to get significance from entangled or loose data and can be utilized to concentrate examples and distinguish patterns that are too unpredictable ever to be seen by either people or other PC methods. These are appropriate for consistent esteemed inputs and yields. For instance, manually written character revamping, for preparing a PC to claim English message and numerous true business issues and have as of now been effectively connected with numerous commercial enterprises. Neural systems are best at recognizing examples or patterns in data and appropriate for expectation or anticipating needs

SNs lacks with the resources and hence it cannot able to perform the huge data processing inside the SNs; but more heterogeneous and huge sensor data demands the perfect, high-level data streaming. These problems in SNs data processing can be solved by using the combination of SNs with the cloud, which process the data in two ways; a) The sensed data by SNs are forwarded to cloud for processing, b) The cloud system will process the query, subdivide the query and forwards to the SNs and responding to the query the sensor node will send only the necessary data.

F. Existing frameworks for Sensor Data Processing (SDP) based on cloud:-

Many authors have used different kind of SDP frameworks based on cloud [8]. One of the framework Cloud view is been designed for storage, analysis and also processing of huge data collected from more number of sensors connected with cloud environment. The combination of both online and offline (also known as hybrid) system adapts the distributed MapReduce and Hadoop for storage and processing the sensor data. The hybrid system possesses the functions in data collection, maintenance, case base creation, updating, features extraction and reduction etc. The Hadoop in the system supports the batch processing style and it will cause problem in real-time processing. Some of the authors are interested in using the Apache Hadoop and MapReduce for large scale sensor data processing (LSSDP).

Many of the authors have proposed a flexible and multi-level SDP system. The SNs, in the system composed of gateway an d coordinator to gather and store the sensor data and also the allows the local storage. Different types of SNs and local servers

(LSs) are interconnected to have computing and backup of local storage. LSs allows the uniform access of stored data in the cloud for the clients, which uses different types of data bases like Tiny DB, Mango DB, SQLite and MySQL. Many authors have used the integrated system of Enterprise Resource Planning (ERP) and SNs, to reduce the resource usage, but the data transmission in SNs may cause the more energy consumption.

G. Considerable poinsts for the designing the SNs Query processing (SNQP) framework:

Following points are needed to be considered while designing the SNQP base on cloud [9].

- The system should allow one time, periodic and continuous query processing.
- The data transmission is needed to be reduced by allowing the combined aggrigation of user defined and system defined.
- Explicit allow of data pre-processing mainly discrepancy or uncertainty management).
- Nearer query processing having exact error tolerance level.
- The system should allow acquisitional query on collected, query processing, and on-demand query processing data.
- The system must allow WSN resource management by using processed data.
- Explicit fault management loads balancing, and data consistency management.
- The system also includes monitoring of power consumption and network legibility problems.

III. RESEARCH WORK

Herodotou et al. [10] a data analytics framework create on Hadoop, concentrates on enhancing the execution of groups all through the data lifecycle in analytics, Without obliging clients to comprehend the accessible con gyration choices. Star sh utilizes systems at a few levels to streamline the execution of Map-Reduce employments. It utilizes dynamic instrumentation to master le occupations and enhances work owes by minimizing the effect of data unbalance and by adjusting the heap of executions.

Tooker et al. [11] examined a framework that has a novel underground correspondence framework and an online underground sensor system proving ground. The framework incorporates an underground receiving wire and underground sensor system proving ground.

Richa et al. [12] has suggested another technique for data mining by actualizing a various leveled virtual k-mean methodology on the profoundly scattered distributed computing. The methodology intends to mine distributed computing data which is topo-graphically circulated in multi-areas and made by various virtual machines. Virtualization methodology is thought to be a key innovation for the advancement of assets and gives adaptability in the data focuses. Data handling lies in this HVKM methodology and it incorporate data spread and data change.

Fan et al. [13] proposed sensor data stockpiling arrangements in light of the Hadoop distributed computing structure. Because of the quick development of sensor data stockpiling and preparing, conventional capacity frameworks are not ready to meet the data access prerequisites.

Pathak et al. [14] talks about the thought of joining remote sensor systems and distributed computing from existing methods. The author presented remote sensor arrange by the virtual sensor in the cloud. The thought is to store the data on both the virtual sensor and genuine sensor.

Zhang et al. [15] examines a portion of the Wireless Sensor Network (WSN) difficulties like the restricted assets of a sensor, constrained battery life, constrained data transmission and constrained preparing force.

Cugola et al. [16] have presented a study that uses the data-driven communication standard received by WSN applications. It likewise considers the qualities (i.e., connection) of the sensors to channel data. Depicts the development of a sensor system utilizing an informing system. An informing system is an overlay system with an arrangement of substance mindful message taking care of capacities. The utilization of an informing system can diminish the many-sided quality and support weight of coordinated sensor data frameworks. Message intercession (an element of the informing system) empowers interoperation of different applications and coordination of various sensor data.

Ghobakhlou et al. [17] talks about the Sensor Network applications in imperative ranges, for example, medicinal services, military, basic base checking, environment observing, and assembling. Because of a percentage of the restrictions of WSNs as far as memory, vitality, processing, correspondence, and versatility, productive administration of the huge number of WSNs data in these territories is a vital issue.

Shmeidman et al. [18] gives an Internet-based base for uniting sensor systems to applications and offers topic based revelation and information handling administrations.

Devi et al. [19] examines novel and appealing answers for data assembling crosswise over transportation, business, medicinal services, mechanical robotization, and natural checking. The up and coming era of WSN will advantage when sensor data is added to writes, virtual groups, and interpersonal organization applications.

Mitton et. al., [20] have presented Cloud based Smart Traffic, a product framework to empower movement data, oversee, break down and present the outcomes in an adaptable, versatile and secure way utilizing a Cloud stage. The foundation is utilized to handle conveyed and parallel data administration and investigation utilizing cosmology database and the well known Map-Reduce structure.

IV. APPLICATION OF SENSORS AND SYSTEMS

An idea of empowering methods such as data analysis, sensor innovation etc., have provided specialists to realize the frameworks in detailed way.

- Applications incorporated with health, home recovery, security, treatment evaluation viability, and scatters pre-identification.
- Monitoring of health and fitness:- helps in remotely monitoring of health and fitness in clinical mediations.
- Safety Monitoring: Many applications were developped for security observation, life alert etc.
- Home Rehabilitation: Rehabilitation, a rising zone of utilization of innovation, is the utilization of adoptable sensors to encourage the usage of home rehabilitation intercessions. Frameworks that expect to facilitate the usage of restoration activity programs frequently influence the blend of sensing technology virtual reality (VR) and intuitive gaming situations.
- Treatment Assessment Efficacy:A quantitative method for evaluating treatment viability can be an important instrument for clinicians in administration. By realizing what happens between outpatient's visits, treatment intercessions can be adjusted to the requirements of individual patients. A additional imperative application would be for utilization in the randomized clinical experiment. By gather target and exact measures of problems.

V. RESEARCH GAP

The identified research gaps of the proposed system are as follows:

- Less focus on computational complexity: Developing a system that can perform analytics on massive sensor data should posse's complexity of data size, data heterogeneity, and data formats (if working on heterogeneous wireless sensor network). Hence, none of the existing literature was found to address this fact about handling the above stated complexities.
- Less focus on benchmarking: Majority of the existing studies are not found to justify their outcomes with respect to comparative analysis of considering uniform test bed.
- Network modelling and topology not emphasized: The cloud topology on which the sensor data is working is never discussed in any literatures. Although the forms of the data being accumulated are just scheduled using TDMA in sensor network, but there are various networking applications especially in healthcare that doesn't work by TDMA scheduling and it works on dynamic type. We didn't come across any study that has addressed this problem.
- Few potential mining algorithms: Although, it is known that mining techniques of conventional network are definitely not application over big data analysis, but still various studies are found to consider conventional mining technique without any justification of their applicability as well as their extensive performance when the dataset changes.

VI. CONCLUSIONS

These papers discuss about the fundamentals of the data being captured from the sensors and are stored and analyzed in cloud. The paper also discusses about live sensor data sustains and clumps of authentic sensor data. It contains recreation clocks and diverse timetables including single, repeating and module driven. Also, the investigation cloud has straightforward programming interfaces, which makes it simple to create examination modules. Plans to move the limits towards a Cloud of sensors and so forth, where sensors and actuators can be found and totaled, as well as powerfully give as an administration, applying the Cloud, provisioning model. Having at the top of the priority list the (concurred) client necessities, it is consequently conceivable to build up Sensors and Actuators as Service suppliers. The Sensing and Actuation as a Service (SAaaS) visualizes new situations and imaginative, universal, quality included applications.Uncovering the detecting and activation world to any client, a client and in the meantime a potential supplier too, in this manner empowering an open commercial center of sensors and actuators. Hence, it is quite clear that sensory applications are quite complicated and therefore the data aggregated from them will be also complex enough to be analyzed over cloud. Hence, our future work will be focussed on developing a prototype that can generate, store and analyze the complex sensor data over cloud.

REFERENCES

[1]. S. R. Wankhade, and N. A. Chavhan, "A review on data collection method with Sink node in wireless sensor network", *International Journal of Distributed and Parallel Systems (IJDPS)* ,Vol.4, No. 1,PP.67-74,2013.

[2]. K. Krishnakumar, , J. Gubbi, and R. Buyya, "A Framework for IoT Sensor Data Analytics and Visualisation in Cloud Computing Environments", *University of Melbourne*,2011.

[3]. F. Wang, and J. Liu, "Networked wireless sensor data collection:issues,challenges,and,approaches", *Communications Surveys & Tutorials, IEEE*,Vol 13, No. 4 ,pp.673-687, 2011.

[4]. *X.* Chen, X. Hu, and J. Zhu, "Minimum Data Aggregation Time Problem in Wireless Sensor Networks", *Springer Berlin Heidelberg*, pp. 133-142., 2005.

[5]. R.S.Ponmagal, M.P.Chitra, P.Dineshkumar, V.N. Rajavarman andG.D.Kumar, "CAEMON: Cloud Access Execution and Monitoring for Big Data Analytics of Sensor System", *ARPN Journal of Engineering and Applied Sciences*, Vol.10, No.2, 2015.

[6]. Y. Ma, M. Richards, M. Ghanem, Y. Guo, and J. Hassard. "Air pollution monitoring and mining based on sensor grid in London", Vol 8, No. 6, 2008.

[7]. A.S. Patil and S. Patil. "A review on Data Mining based Cloud Computing", *International Journal of Research in Science & Engineering*,Vol 1,No 1, 2010.

[8]. P.Bihani and S. T. Patil, "A Comparative Study of Data Analysis Techniques", *International Journal of Emerging Trends & Technology in Computer Science (IJETTCS)*,Vol 3, No 2, PP. 95-101,March – April 2014.

[9]. C.R. Chowdhury "A Survey on Cloud Sensor Integration", *International Journal of Innovative Research in Computer and Communication Engineering*, Vol. 2, No. 8, August 2014.

[10]. H. Herodotou, H. Lim, G. Luo, N. Borisov, L. Dong, F. B. Cetin, and S. Babu, "Starfish: A Self-tuning System for Big Data Analytics." *In CIDR*, Vol. 11, PP. 261-272, 2011.

[11]. J. Tooker, X. Dong, M. Vuran, and S. Irmak, "Connecting Soil to the Cloud: A Wireless Underground Sensor Network Testbed", *In Sensor, Mesh and Ad Hoc Communications and Networks (SECON), 9th Annual IEEE Communications Society Conference on*, PP. 79-81, 2012.

[12]. Richa and S. Mittal ,"Data Mining Approach IDS K-Mean using Weka Environment", *International Journal of Advanced Research in Computer Science and Software Engineering*, Vol. 4, Issue. 8, pp. 482-488, 2014.

[13]. T. Fan, X. Zhang, and F. Gao, "Cloud Storage Solution for WSN in Internet Innovation Union", *International Journal of Database Theory and Application*, Vol.6, No. 3, pp.49-58, 2013.

[14]. M. Pathak, "An Approach to Memory management in Wireless Sensor Networks", *International Journal of Computer Science & Engineering Technology*, Vol. 4, No. 08, Aug 2013.

[15]. P. Zhang, Z. Yan, and H. Sun. "A novel architecture based on cloud computing for wireless sensor network", *In Proceedings of the 2nd International Conference on Computer Science and Electronics Engineering*, 2013.

[16]. G. Cugola, and M. Migliavacca. "A context and content-based routing protocol for mobile sensor networks." *In Wireless Sensor Networks, pp. 69-85. Springer Berlin Heidelberg*, 2009.

[17]. A. Ghobakhlou, A. Kmoch, and P. Sallis, "Integration of Wireless Sensor Network and Web Services", *In Proceedings of the 20th International Congress on Modeling and Simulation*, Vol. 16, pp.838-844, 2013.

[18]. J. Shneidman, P. Pietzuch, J. Ledlie, M. Roussopoulos, M. Seltzer, and M. Welsh. "Hourglass: An infrastructure for connecting sensor networks and applications". *Cambridge, MA: Harvard technical report*, 2004.

[19]. S.J. Devi, G. S.Devi,andG. M. T.Selvan,"Wireless Sensor Network Integrating with Cloud Computing for Patient Monitoring", *International Journal of Engineering Sciences & Emerging Technologies*, Vol. 6, No. 3, pp.316-323, 2013.

[20]. N. Mitton, S. Papavassiliou, A. Puliafito, and K.S. Trivedi, "Combining Cloud and sensors in a smart city environment", *EURASIP journal on Wireless Communications and Networking*, No. 1, pp.1-10, 2012.